MELBOURNE INSTITUTE
Applied Economic & Social Research

Working Paper Series
Students Are Almost as Effective as Professors in University Teaching

Jan Feld
Nicolas Salamanca
Ulf Zölitz

Working Paper No. 23/17
September 2017

THE UNIVERSITY OF MELBOURNE

MELBOURNE INSTITUTE

Students Are Almost as Effective as Professors in University Teaching*

Jan Feld[†], Nicolas Salamanca[‡] and Ulf Zölitz[§]
[†] School of Economics and Finance, Victoria University of Wellington; and IZA Institute of Labor Economics
[‡] Melbourne Institute: Applied Economic & Social Research, The University of Melbourne; and IZA Institute of Labor Economics
[§] Behavior and Inequality Research Institute (briq); IZA Institute of Labor Economics; and Department of Economics, Maastricht University

Melbourne Institute Working Paper No. 23/17
September 2017

* We would like to thank Luke Chu, Julie Moschion, Cain Polidano, Chris Ryan, Ingo Isphording, Derek Stemple and Barbara Wolfe for helpful comments. We further thank Sylvia Beenen, Joël Castermans, Jeanette Holmes, Paul Jacobs, Sylvie Kersten, Sanne Klasen, Kim Schippers and the SBE Educational Research and Development Department for helpful discussions, valuable background information and data. We thank Sophia Wagner for providing outstanding research assistance. This research was partly supported by the Australian Research Council Centre of Excellence for Children and Families over the Life Course (project number CE140100027). The Centre is administered by the Institute for Social Science Research at The University of Queensland, with nodes at The University of Western Australia, The University of Melbourne and The University of Sydney. Corresponding author: Ulf Zölitz. briq – Behavior and Inequality Research Institute GmbH, Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany. Phone: +49 (228) 3894-703; Fax: +49 (228) 3894-510; Email: ulf.zoelitz@ briq-institute.org.

Melbourne Institute: Applied Economic & Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* +61 3 8344 2100**
***Fax* +61 3 8344 2111**
***Email* melb-inst@unimelb.edu.au**
***WWW Address* melbourneinstitute.unimelb.edu.au**

Abstract

Many universities around the world rely on student instructors—current bachelor's and master's degree students—for tutorial teaching, yet we know nothing about their effectiveness. In a setting with random assignment of instructors to students, we show that student instructors are almost as effective as senior instructors at improving their students' short- and longer-run academic achievement and labor market outcomes. We find little heterogeneity across different course types, student characteristics, or instructors' personal academic quality. Our results suggest that the use of student instructors can serve as an effective tool for universities to reduce their costs with negligible negative effects on students.

JEL classification: I21, I24, J24

Keywords: Student instructors, university, teacher performance

1 Introduction

In many universities worldwide, student instructors are an integral part of university teaching.[1] These instructors are currently enrolled bachelor's and master's students who typically teach small groups of students in the tutorials, exercise sessions, or lab sessions that complement course lectures. From a university standpoint, the main advantage of using student instructors is that they are substantially cheaper than more senior staff.[2] However, student instructors are also wildly inexperienced and much less qualified. It therefore appears questionable whether they can provide the same quality of education and prepare students for the labor market as effectively as professors or lecturers.

This paper investigates how student instructors affect their students' academic performance and labor market outcomes. We use data from Maastricht University's School of Business and Economics (SBE), an institution with two key features that make it ideal for studying the effect of student instructors. First, in many courses student instructors teach tutorials side by side with more senior staff using identical course material, providing the necessary within-course variation in teacher type. Second, student assignment to tutorial groups—and therefore to instructors—is random, conditional on scheduling constraints. This allows us to estimate the effect of instructor type on student outcomes without worrying about endogenous matching of students and instructors.

We find that, on average, being assigned to a student instructor reduces students' grades by 2.3 percent of a standard deviation, a small and only marginally significant effect. This effect seems to be driven by a slightly disproportionate negative effect of student instructors

[1] Surveying people with experience in higher education in the OECD, we find that student instructors are used in 26 out of 35 OECD countries (see overview Table A1 in the appendix). While undergraduate teaching assistants are less prevalent in US, Teaching Assistants (TAs) including masters' and PhD students currently account for about 11.4 percent of the total employment of postsecondary teachers in the US (Bureau of Labor Statistics (OES) report, 2015).

[2] Median wages for all types of post-secondary teachers in the US are $68,010, while median wages of TAs are only $32,490 (Bureau of Labor Statistics (OES) report, 2015).

on lower ability students. We do not find any detectable impact of student instructors on student grades in subsequent courses. Since the point estimates on current and subsequent grades are precisely estimated, we can rule out even modest-sized effects of student instructors. When looking at students' course evaluations, we find only weak evidence that student instructors are evaluated more negatively. Finally, we find no measurable impact of student instructors on students' longer-run outcomes, such as job search length after graduation, earnings, job satisfaction or retrospective study satisfaction.

To date, the question of whether student instructors are as effective as other university instructors has been neglected. There are, however, a few related papers which study how the origin and ethnicity of graduate teaching assistants (TAs) affect student performance. Lusher, Campbell, and Carrell (2015) study the role of graduate TAs' ethnicity and find that students' grades increase when they are assigned to same-ethnicity graduate TAs. Borjas (2000) and Fleisher, Hishimoto and Weinberg (2002) study the effect of foreign-born graduate TAs, and reach opposing conclusions. While Borjas (2000) finds that foreign-born TAs negatively affect student grades, Fleisher et al. (2002) find that foreign-born graduate TAs have negligible effects on student grades and that, in some circumstances, these effects can even be positive. Bettinger, Long and Taylor (2016) look at the effect of having a PhD student as a full instructor (rather than a TA) on students' subsequent major choices. They find that students are more likely to major in a subject if the first courses in that subject are taught by a PhD student. None of these studies, however, compares the effectiveness of student and non-student instructors.

Our results are consistent with the idea that student instructors do not differ in their value added when compared to senior instructors. In the teacher value added literature, Rivkin et al. (2005) show that, while teachers significantly affect achievement, little of the variation in teacher quality can be explained by observable teacher characteristics such as their education or experience. Chetty et al. (2014a, 2014b) show that students exposed to high value-added teachers in primary school have a higher probability of attending college, attending higher-quality colleges and having higher earnings over their life cycle. Another related strand of literature looks at the effect of instructor characteristics on student outcomes at the university level. Bettinger and Long (2010) and Figlio, Shapiro and Soter (2015) find a positive effect of adjunct instructors compared to tenure track and tenured instructors on student performance. Hoffmann and Oreopoulus (2009) find that objective instructor characteristics, such as academic rank and salary, do not predict student performance, yet students' evaluations of their teachers are positively correlated with student performance. De Vlieger, Jacob and Stange (2017) find that instructor performance in a college algebra course at a large for-profit university grows modestly with course-specific teaching experience, but is unrelated to pay. Fairlie, Hoffmann and Oreopoulos (2014) find that minority students benefit from minority instructors. Yet, again, none of these literatures explicitly study student instructors despite their importance in higher education.

Student instructors are quite different from any other type of instructor in a university setting. The most apparent difference is the immense gap in qualifications and experience, which suggests that student instructors would, in principle, be less effective than more senior instructors. However, student instructors could be better able to relate to students through shared characteristics and experiences, which could give them an advantage to teach more effectively. From the university's perspective, student instructors are inexpensive in terms of direct and indirect remuneration and in terms of the overhead they require. Moreover, there is a constantly-renewing pool of student instructors that requires little additional recruiting efforts, since they are often former course-takers. With such stark differences in instructor type, and given their extensive use in university teaching worldwide, it is crucial to assess the impact of this unique low-cost teaching resource. In the remainder of the paper, we describe in detail how we quantify and characterize their performance.

2 Institutional Background and Data

2.1 Institutional Environment

To estimate the effect of student instructors on student outcomes, we use data from Maastricht University's School of Business and Economics (SBE) from the academic years 2009/2010 to 2014/2015.[3] The bulk of teaching at the SBE is done in four regular teaching semesters of eight weeks each, where students typically take two courses simultaneously.[4] Over the entire eight-week teaching semester, students usually receive three to seven lectures for each course. The bulk of the teaching, however, is done over two-hour tutorials which occur twice per week for each course. These tutorials are at the center of our analysis. The tutorials are organized in groups of up to 16 students who are assigned to one instructor—either a student instructor or a more senior one. In these tutorials, students discuss the course material and are guided by the instructor, who is always present. While this discussion-based teaching style is used in several universities at the postgraduate level, the SBE stands out by also using it at the undergraduate level. Tutorials are crucial at the SBE: attendance is compulsory and recorded by instructors, tutorial participation and attendance is often graded, non-attendance can easily result in failing the course, and the SBE guidelines explicitly prohibit switching between assigned tutorial groups. Within a given course, tutorials are also quite homogeneous: they use identical course material, they have the same assigned readings and exercise questions, and they follow the same course plan.

In many courses, tutorial groups are taught by a mixture of student instructors and more senior instructors. This within-course variation in instructor type is our source of identifying variation. We estimate the effectiveness of student instructors compared to senior instructors,

---

[3] For more detailed information on the institutional environment see Feld and Zölitz (2017) and Zölitz and Feld (2016).

[4] We use 'course' throughout to refer to a subject-year-period combination. Thus, we consider, e.g., Microeconomics in period 1 of 2011 and Microeconomics in period 1 of 2012 as two separate courses.

which include post-docs, lecturers, assistant professors, associate professors and full professors.[5] Our setting thus allows us to obtain local treatment effect estimates for tutorials in courses that use both student and non-student instructors, which identify the effect of increasing the usage of student instructors in courses that already use them. These estimates can thus inform school level policy on the adjustment of the intensive margin of student instructor use.

Student instructors are typically recruited by an SBE education manager and approved by the course coordinators. The most important characteristics in the recruitment process are the students' grades, previous experience with the course at hand, and a sufficient command of English, the language of instruction for all courses. Next to their low academic rank, student instructors stand out because they lack teaching experience. In the six-year period covered by our data, student instructors teach on average 1.9 courses, compared to the average of 3.2 and 5.8 courses taught by PhD students and senior instructors.

It is much more inexpensive for the SBE to employ student instructors than any other staff type. On a per-hour, within tutorial group basis, and ignoring overhead cost differences, student instructors are four times less expensive than a newly hired assistant professor and five times less expensive than full professors in the lowest salary scale.[6] The search and hiring costs of student instructors are also close to zero. They can easily be recruited from the constantly-renewing pool of students taking each course, and they are offered standard short-term contracts, often as short as a couple of weeks in order to cover teaching staff gaps. Thus, student instructors represent an elastic, convenient, and low-cost labor force for the university.

[5] We treat PhD instructors as a separate sub-group, but since they are not the focus of this paper we do not explicitly report these results.
[6] Calculation based on teaching loads common for SBE employees. For more information regarding salary scales see:
https://www.maastrichtuniversity.nl/support/um-employees/money-matters/salary-payment-and-statement.

2.2 *Summary Statistics*

To estimate the effect of student instructors, we limit our estimation sample to courses that have at least one student instructor and one non-student instructor.[7] Table 1 provides an overview of the courses offered by SBE and the courses that are part of our estimation sample. Table 1 shows that student instructors are disproportionally used in large bachelor-level courses. This likely reflects the larger need for teaching staff in these courses. Interestingly, we do not find significant differences in the use of student teachers in courses by the average grade point average (GPA) of the students enrolled or by whether the course is mathematical or non-mathematical (as defined in Section 3.3). This indicates that student instructors are *not* systematically allocated to simpler, non-mathematically demanding courses, as one might have suspected given their lack of formal qualification and experience.

Table 2 presents summary statistics aggregated at the instructor level (Panel A), at the student level (Panel B), and at the tutorial group level (Panel C) in our estimation sample. We observe 434 instructors who teach 6,649 different students in a total of 2,534 different tutorial groups. Half of all instructors are student instructors, and they teach 42 percent of all the tutorial groups in our sample. For comparison, PhD students represent 25 percent of all instructors, teaching 18 percent of tutorial groups, while senior staff account for 24 percent of the instructors and 40 percent of the taught tutorial groups in our estimation sample. Instructors' nationalities at the SBE are quite diverse, with the single largest nationalities being German (43 percent) and Dutch (30 percent) and the rest coming from various other countries. About 38 percent of instructors are female.

The student performance data in our main estimation sample consists of 28,203 course final grades for the 6,649 students in our sample. The final course grade usually consists of

[7] Our core dataset has information on 103,664 course final grades from 14,089 students who took 1,354 courses, taught by 772 instructors over 24 teaching periods between 2009/2010 and 2014/2015. However, we make some restrictions on our core data which affect all tables in this paper. See Appendix A1 for details.

multiple graded components, with the highest weight typically placed on the final exam. The components and weights, however, vary from course to course. Some of the components of the final grade, such as group work or tutorial participation, are directly graded by the students' own instructor. In our data, we only observe course final grades. Differences in grading standards between instructor types can therefore partially drive part of our estimates, a concern we address in Section 3.3. The Dutch grading scale ranges from 1 to 10, with 5.5 as the lowest passing grade. Figure 1 shows that the distribution of the course final grades covers the entire range of possible grades, providing us with sufficient variation in our measure of academic performance. Throughout our analyses, we account for differences in student ability using students' GPA, constructed as the average of all grades prior to the current course, weighted by course credit points.

2.3 Assignment of Instructors and Students to Tutorial Groups

Both students and instructors are assigned to tutorial groups within each course in a manner that results in random assignment of students to either a student instructor or a more senior instructor. In the scheduling process, students are first randomly assigned to tutorial groups conditional on scheduling conflicts.[8] For all bachelor students, this assignment was unconditionally random until the academic year 2009/2010. From 2010/2011 onwards the schedulers balanced tutorial groups by nationality (making sure that the proportion of German, Dutch, and other nationality students were the same across tutorial groups in each course), but otherwise the assignment remained random. In previous work with data from the same

[8] Courses are usually scheduled in a way to avoid scheduling conflicts. For example, the first-year compulsory courses that students take in parallel are scheduled on different days. The main source of scheduling conflicts is students taking different elective courses. To account for potentially non-random assignment due to other courses taken at the same time, we control for fixed effects for all combinations of courses that students take in each period. A small number of students have other scheduling conflicts because they take language courses, work as student instructors, have regular medical appointments, or are top athletes and need to accommodate inflexible training schedules. Importantly, none of these exceptions is a response to the instructor or students of a tutorial group.

environment, we show that tutorial group assignment has the properties we would expect under random assignment (see Feld & Zölitz, 2017, and Zölitz & Feld, 2016). Instructors are then assigned to tutorial groups, generally in consecutive time slots, and, importantly, this assignment is unrelated to student characteristics. About 10% of instructors in each period indicate time slots in which they are not available for teaching. However, this happens prior to any scheduling of students or other instructors, and requires the approval of the department chair. Taken together, neither students nor instructors nor course coordinators influence whether a student is assigned to a student instructor or a more senior instructor.

Random assignment of students to tutorial groups implies that instructor characteristics are, on average, unrelated to observable and unobservable student characteristics. To support this claim, we test whether in our estimation sample instructor type is related to four 'pre-assignment' student characteristics: previous GPA, gender, age, and the rank of the student ID—a proxy for tenure at the university. We do this by regressing each of these four pre-assignment characteristics on student instructor and PhD student instructor dummies (keeping senior instructors as the base group), including fixed effects for all course and parallel course combinations as well as time-of-the-day and day-of-the-week fixed effects as controls.

Table 3 shows the results of these balancing tests. Columns (1), (3) and (4) show that instructor type is not significantly related to students' GPA, age, and ID ranks. Column (2) shows that student instructors are marginally less likely to teach female students. However, the size of this difference, a 1.1 share in female students, is tiny. Moreover, only one out of eight coefficients of interest we tested is statistically significant, and any method to account for multiple testing eliminates this significance. Finally, a joint F-test of the student instructor and PhD student dummies, which tests for overall differences in assignment to different instructor types, cannot reject the null of no differences (p-value = 0.209). We nevertheless account for these small differences in gender composition by including student gender dummies in all specifications throughout the remainder of the paper.

Overall, our results confirm that instructor assignment is not systematically related to student characteristics. This puts confidence on the random nature of instructor assignment to students and implies that we can estimate the causal effect of student instructors via least-squares regressions without worrying about endogenous matching of instructors and students.

3 The Effect of Student Instructors on Student University Performance

3.1 Empirical Strategy

We estimate the effects of student instructors on student outcomes via variations of the following model

$$y_{ic} = \beta_1 \text{student instructor}_{ic} + \beta_2 \text{PhD}_{ic} + \gamma' X_{ic} + \delta_c + \varepsilon_{ic}, \tag{1}$$

where y_{ic} is the outcome of student i in course c, and the main regressor of interest is $\text{student instructor}_{ic}$, an indicator of whether student i in course c was taught by a student instructor. We control for PhD_{ic}, an indicator of whether the instructor is a PhD student instructor, which leaves senior instructors (post-docs, lecturers, and assistant, associate and full professors) as the base group. The vector X_{ic} includes several control variables which we vary across specifications and which can include: student gender, student nationality, a cubic polynomial in student age, and the student's GPA before taking the course. The term δ_c represents course-invariant unobserved heterogeneity, which can include factors systematically related to student selection into courses, and other idiosyncratic course characteristics such as the student composition of each course. We account for this heterogeneity by including a complete set of course and other-course combination fixed effects (effectively capturing all possible course combinations taken by students in each period), together with time-of-the-day and day-of-the-week fixed effects for the tutorial group's timing. These fixed effects eliminate any endogeneity from the non-random assignment of students to tutorials that could have stemmed from the parallel course that the student is taking at the same time and restricts our estimates to be identified solely through within-course variation. Finally, ε_{ic} is an idiosyncratic error term in the student outcome process, which is assumed to be uncorrelated with all the regressors and with δ_c. We cluster the standard errors at the instructor level to conduct inference allowing for correlations in student outcomes within an instructor. We standardize our main dependent variable, course final grades, to have an overall mean of zero and standard deviation of one across our estimation sample to make the results easier to interpret.[9]

3.2 Effect of Student Instructors on Course Grades

Table 4 shows the estimates of the effect of student instructors on final grades with different sets of controls. As expected under random assignment, the point estimates are similar across all specifications. The estimated effect in our most parameterized specification in Column (3) is negative, small, and marginally statistically significant. The point estimate indicates that having a student instructor instead of a more senior instructor decreases a student's grade by 2.3 percent of a standard deviation. The effect size is similar in magnitude to estimates from Lusher et al. (2015), who find that Asian students' grades increase by 2.3 percent and non-Asian student grades increase by 3.7 percent when exposed to TAs of their own ethnicity. In terms of the Dutch 1-to-10 grade scale, the effect size is equivalent to a reduction of 0.04 points.[10] To place the size of this effect in perspective, this is less than the average performance gap between the median and the 51st percentile student in terms of student ability. This effect

[9] As a pre-analysis check, we first test whether there are meaningful differences in instructor effectiveness, which is an essential precondition for analyzing the performance of different instructor sub-groups. To do this, we estimate a version of the model in Equation (1) where we replace instructor type with instructor fixed effects. The joint F-test of instructor fixed effects (Baltagi, 2005, p.13) rejects that all instructors affect students' grades equally (p-value < 0.015, see also Figure A1 in Appendix). This confirms that there are differences in instructor effectiveness and provides a good starting point for testing if student instructors differentially affect student outcomes.

[10] We have also tested whether student instructors affect the probability of dropping out of the course. We find no evidence of that, which is not surprising given that the dropout rate in our estimation sample is only 4 percent.

is also dwarfed by other determinants of student grades in the same environment, such as the 0.17 standard deviation grade premium received by students with the same nationality as their graders (Feld et al., 2016).[11] These results show that student instructors have an economically insignificant effect on student performance.

We find the lack of an effect of student instructors on grades surprising given their lack of formal qualification and experience, especially since it is generally believed that student grades benefit from more experienced and better qualified teachers. However, the recent review in Harris and Sass (2011) concludes that in primary and secondary education the effects of teachers' qualifications and experience on student performance are mixed and that there is no consensus in the literature. The small effects reported in Table 4 are thus surprising, yet not unheard of, and they prompt us to further explore why student instructors hardly affect student performance.

3.3 Grading Standard, Math Courses and Student Ability

Due to their lack of experience or their closeness to the students they are in charge of teaching, there may be some factors, specifically in terms of their teaching and grading methods, that could affect our results. For example, we might be worried that the overall small effects are a result of student instructors grading more generously, thus cancelling out any detectable penalties on the performance of their students. Student instructors may, for example, want to compensate the students by giving them higher participation grades. Within our institutional setting, we can explore if this is the case by estimating the effect of student instructors for first-year courses and non-first-year courses (i.e., second year, third year, and master's courses) separately. In first-year courses, instructors have a negligible influence on the grading standard

[11] Our main results are robust to the inclusion of gender and nationality matches between instructors and students, which may be another source of grading bias. See Feld et al. (2016) for further discussions on grading biases at the SBE and for a more detailed explanation of the examination procedure.

since the final grade consists entirely of the final exam grade. These final exams are largely machine-graded multiple-choice questions.[12] If the effects are small because of compensating grading biases, student instructors should have a more negative impact on student grades in first-year courses.

Columns (1) and Column (2) of Table 5 report regression results for first-year and non-first-year courses separately. The estimated effect of student instructors in first-year courses is even closer to zero than the estimated effect for the whole sample, whereas in non-first-year courses having a student instructor reduces students' grades by a slightly larger 3.3 percent of a standard deviation. These point estimates may indicate that student instructors grade *less* generously, or they may be the result of other differences in the role of instructors between early- and late-program courses. Importantly, these results are evidence against the concern that our main effects are small because student instructors grade more generously.

In Columns (3) and (4) of Table 5 we report separate regressions for mathematical and non-mathematical courses. In non-mathematical courses, students score 3 percent of a standard deviation lower if they have a student instead of a senior instructor. In mathematical courses, we find virtually no difference between the effectiveness of student and senior instructors. The results are consistent with the idea that student instructors are better at teaching narrowly-defined course material, but that they lack the experience or broader knowledge to effectively teach less technical courses.

Finally, we ask whether the effect of student instructors is smaller for students who can independently understand the course material. In this case, we would expect student instructors to matter more for less able students. To test this hypothesis, we use students' GPA prior to enrollment in each course as a measure of ability and categorize each student as *lower ability*

[12] While some instructors help out with the grading of non-machine graded part of exams, they usually mark the same question for all students in the course so that potential differences in instructors' grading standards affect students of all tutorial groups equally.

if their GPA is in the bottom half of the course-specific GPA distribution and *higher ability* otherwise.

In Columns (5) and (6) of Table 5 we show separate regressions for lower- and higher-ability students. Student instructors are more detrimental for lower ability students, while higher ability students appear to be unaffected by instructor type. Lower ability students' grades decrease by 4.5 percent of a standard deviation when they are exposed to student instructors, a statistically significant but still small effect. These point estimates are consistent with our hypothesis that student instructors can be more harmful to less able students.

3.4 Effect of Student Instructors Across the Student Grade Distribution

We now turn to the question of how the effect of student instructors differs at different parts of the student grade distribution. Student instructors may, for example, have a stronger effect on students who are at the margin of passing the course, in which case their impact on student outcomes will be understated when just looking at the average student grade. Our result for lower ability students leaves this possibility open but is too crude to detect any effect at this important margin. We therefore estimate the effect of student instructors at each point in our discrete grade distribution using an adaptation of the unconditional quantile regression in Firpo, Fortin, and Lemieux (2009).[13] These estimates can be interpreted as the impact of having a student instructor on student grades at *each point in the final grade distribution*, holding constant other characteristics.

Figure 2 shows the impact of student instructors at each point in the course grade distribution (see also Table A2 in the Appendix). The overall pattern indicates that across the largest and most densely populated part of the grade distribution student instructors have an

[13] Our method calculates their Recentered Influence Function (RIF) for each of the points in our discrete grade distribution (i.e., at final grade = 1.5, 2, …9.5, 10), replacing their kernel estimator for the density at each point in their continuous outcome distribution, $\widehat{f_Y}(q_\tau)$, by the corresponding probability mass point in our discrete outcome distribution.

equally negligible effect on student grades. This is consistent with the small average effects of student instructors presented in the sections above. The figure also suggests that student instructors can be detrimental for students at the bottom of the grade distribution (i.e., for students who are already failing the course), although these effects are imprecisely estimated. Most importantly, we do not find an effect around the minimal passing grade threshold of 5.5, indicating that student instructors do not affect their students at this crucial margin.

3.5 Cumulative Effects on Course Grades

While we have shown that having *one* student instructor in any given course has a negligible effect on grades, the effect of multiple student instructors could add up. The existence of dynamic complementarities in human capital formation (Cunha and Heckman, 2007) opens the possibility of cumulative and non-linear effects of student instructors. Depending on the dynamic structure of university learning, it could well be that we fail to detect negative effects of student instructors at the mean but that being exposed to several student instructors can eventually be detrimental to student performance.

To test for potential cumulative effects, we estimate a version of Equation (1) where we interact the student instructor variable with the number of previous student instructors each student has been exposed to. Students in our data differ widely in the amount of student instructors they have been exposed to, with almost 20 percent of our sample being exposed to more than three student instructors. Figure 3 shows the estimated effect of student instructors from this regression (see also Table A3 in the Appendix). The effects show no obvious pattern, with all point estimates being small and invariant regardless of the number of previous student instructors, and a formal F-Test failing to reject the null of equal effects (p-value=0.790). Students who have been exposed to five or more student instructors in the past seem to be slightly affected by an additional student instructor, but these events are relatively infrequent and thus the estimates are not precise enough to allow us to draw strong conclusions. Overall, we find no evidence of cumulative effects of student instructors.

3.6 *Effects of Student Instructors in Subsequent Course Grades*

One may be concerned that student instructors affect learning in a way that is not reflected in current grades. Student instructors may, for example, teach more to the test, while senior instructors may help the students get a deeper understanding of the course material. If this is the case, we would expect that, even if there is no measurable difference between student and senior instructors in current grades, student instructors affect students' grades in *subsequent* courses, i.e., grades after the students have been taught by student instructors. Effects on future grades would identify persistent effects, if any, of student instructors on student performance, a concept closely related to the effects of teachers on 'deep learning' (Carrell and West, 2010, p. 412).

We estimate the effect of student instructors on subsequent grades for a limited sample of bachelor students whom we observe for all three years of their program. In the SBE, bachelor programs begin with a set of program-specific first year compulsory courses. From the second year onwards, students can choose some of the courses they take, and by the end of the second year students need to commit to a major within each program since they need to start taking their major-specific courses. Throughout this whole process, instructors remain randomly assigned to students. We measure the effect of exposure to a student instructor during the first-year compulsory courses on the average grades obtained in second- and third-year courses. The corresponding econometric model can be expressed as:

$$\bar{y}_{i2-3} = \tilde{\beta}_1 \text{student instructor}_{ic} + \tilde{\beta}_2 \text{PhD}_{ic} + \tilde{\gamma}' X_{ic} + \tilde{\delta}_c + \epsilon_{ic}, \qquad (2)$$

where the only difference to Equation (1) is the dependent variable $\bar{y}_{i2-3}$, which is the average grade of all second and third year courses of student i. The student instructor coefficient, $\tilde{\beta}_1$, reflects both the effect of student instructors on student's subsequent grades which will also be partly driven by any possible effect that student instructors might have on student's subsequent course choices.

Table 6 shows estimates on subsequent grades with different sets of control variables. The coefficient of student instructor is again remarkably stable across specifications and shows no statistically significant effect in any of them. The point estimates are tiny and as precisely-estimated as our main estimates. These results indicate no measurable effects of student instructors on student deep learning and suggest that any small effect of student instructors on grades does not carry over to subsequent courses.

3.7 Heterogeneous Effects by Student Instructor's Academic Ability

Given our findings that student instructors can deliver similar educational quality for a fraction of the costs, one may wonder what would happen if the number of student instructors increased. If the quality of each additional student instructor decreases—as would naturally occur if universities are recruiting the best available students as tutors—it is possible that such a policy would affect student performance. Jespen and Rivkin (2009) show this to be a valid concern by analyzing a class size reduction policy in California that resulted in many lower-quality teachers being hired. They find that the positive effect of a reduction in class size is partly offset by a decrease in teacher quality. At the SBE, the main criteria for hiring student instructors is their GPA. Increasing the number of student instructors would therefore likely mean that the SBE would hire students with lower GPAs.

We can explore the differential effects of student instructors' academic ability since we observe the first-year GPA of 89 out of 217 of our student instructors.[14] Figure 4 shows the

[14] We do not observe the first-year GPA of student instructors who did their undergraduate at other universities or who took their first-year courses at the SBE before the 2009/2010 academic year.

distribution of student instructors' GPAs, revealing that student instructors do have higher grades than the average student, yet there is substantial heterogeneity in their grades and therefore in our measure of instructor academic ability.

Column (1) of Table 7 shows estimates of the effect of student instructors' standardized GPAs on student grades. The estimated effects suggest that instructors with a higher GPA increase their student performance further; however, these effects are not statistically significant. We further estimate the effect of instructor ability non-parametrically by including indicators for above- and below-median-GPA instructors (Column 2) and by including dummies for five quintiles of student instructor GPA (Column 3). Student instructor GPA is not statistically significant in either of these specifications. Furthermore, the small differences between point estimates and the negligible increase in goodness of fit of these models compared to our baseline estimates in Table 4 show that an instructor's GPA is a poor predictor of their students' academic performance.

We cannot rule out that any additional student instructors hired at the SBE would not differ from the incumbent student instructors in other ways not captured by GPA. They may, for example, be less motivated to teach. However, we view these results as suggestive evidence that hiring additional student instructors with lower academic ability is unlikely to lead to worse student outcomes.

4 The Effect of Student Instructors on Course Evaluations and Student Labor Market Outcomes

4.1 Effects on Students' Course Evaluations

Even though student instructors only have a small effect on grades, they may well affect other aspects of students' experiences at the university. The negligible effect on grades may, for example, be a result of students compensating for the low instructional quality of student instructors by studying more outside the classroom. More generally, it could be that student instructors decrease the non-pecuniary benefits of education for their students. If this is the case, increasing the number of student instructors would impose a cost on students that we do not captured by only looking at grades. To explore these issues, we use the extensive individual level student course evaluations at the SBE. These course evaluations ask several questions to students at the end of the period of instruction but before they take their final exams.[15] These data allow us to peek inside the 'black box' and explore several other facets of (perceived) instructional quality and students' reported study effort.

We first estimate the effect of student instructors on four important outcomes reported by students: the overall instructor rating, whether the instructor encouraged group discussion (as is often required at the SBE), whether the instructor stimulated knowledge transfer to other contexts, and whether the instructor mastered the course content.[16] Figure 5 shows the student instructor coefficient of regressions on each of these outcomes with the usual covariates. Our results suggest that student instructors are perceived as significantly worse at transferring knowledge to other contexts and worse at mastering the course content. Both effects are not surprising given their lack of experience and qualifications. The estimated effects on overall evaluation and encouragement of group work are also negative, although not statistically significant.

We then estimate the effect of student instructors on four other student-rated outcomes which, while not directly related to instructor performance, could be affected by it: the overall course rating, the rating of the tutorial group functioning, the rating of the course material, and the self-reported student study hours. Figure 6 shows that having a student instructor leads to significantly worse evaluations of the course overall, which suggests that the student

[15] See Feld and Zölitz (2017) for more detailed description of the course evaluation procedure at the SBE. The average response rate for course evaluation surveys is 38% in our estimation sample. Table A4 in the Appendix shows that questionnaire response is unrelated to instructor type.
[16] See Table A5 in the Appendix for summary statistics of all the questions pertaining to this subsection in our estimation sample and the full text of each question.

experience in the course is less enjoyable. Interestingly, students who are exposed to student instructors also rate the course material as less helpful. This effect suggests some complementarities between instructional quality and the course material, which students themselves might not be able to distinguish when rating the different course components. Importantly, student instructors do not seem to affect the effort provision of their students, though the point estimate suggests that students do spend *less* hours studying when exposed to student instructors. This is, if anything, evidence against students' compensating for low instruction quality by exerting higher effort. Overall, the results in this section suggest that student instructors are perceived as being of lower quality in a few facets by their students. [17]

4.2 Effects on Labor Market Outcomes

Despite having little impact on student performance, student instructors may negatively affect students' labor market outcomes in the longer term. Student instructors might, for example, be less able to provide their students with the skills, knowledge or referrals necessary for beginning a successful career after graduation. Moreover, if student instructors and the education they provide is generally perceived as less worthwhile (see the previous section), they could discourage further human capital investment and negatively impact their students' earnings, employment, and job satisfaction.

To estimate the effect on labor market outcomes, we use an SBE graduate survey which includes data for 1,618 students from our estimation sample who graduated between September 2010 and September 2015.[18] The survey includes questions about job search length after

[17] See Table A6 in the Appendix for the regressions corresponding to Figures 5 and 6.
[18] We conducted the survey in cooperation with the SBE Alumni Office that provided us with contact details for 4,215 out of the 5,504 bachelor students in our estimation sample. We first contacted the graduates via email and provided them with a link to the online survey. We then hired a team of current SBE students who called the graduates who did not respond to the online survey to conduct the survey over the phone. Out of the contacted graduates, 1,618 responded to either the email or phone survey, which means that we have labor market outcome information for 29.39 percent of bachelor students in our estimation sample. Panel A of Table A7 in the Appendix compares our estimation sample with the overall sampling population for this survey.

graduation, earnings in the first job after graduation, current earnings and job satisfaction as well as retrospective study satisfaction. This survey allows us to link university data with labor market outcomes, which are typically not available in other existing studies on the impact of university instructors.[19]

Table 8 shows results from regressions of survey outcomes on instructor type and the same set of covariates we included in the previous regressions. Having a student instructor is associated with a 0.2 percentage point higher probability of having found a job by the time of graduation (Column 1) and a reduction of job search length of 2.25 days (0.075*30 days, Column 2). These point estimates are small and insignificant, yet precisely estimated, which means that we can rule out even modest negative effects of student instructors on job search outcome after graduation.

Our results on earnings are less conclusive. Columns (3) and (4) show that having a student instructor is associated with a 1.7 percent decrease in starting salary and a 1.9 percent decrease in current salary. While these estimates are not statistically significant, they are too imprecise to rule out economically important effects. Columns (5) and (6) shows that assignment to student instructors does not significantly predict retrospective satisfaction with studies or job satisfaction, with small estimates for both outcomes. Overall, we do not detect any evidence that student instructors affect their students' subsequent labor market outcomes in any meaningful way.

4.3 Increasing Power and Correcting for Multiple Testing

The fact that our analyses in the two subsections above are based on 14 different outcomes causes two related problems, which we address in this subsection. The first is a problem of

[19] Panel B of Table A7 in the Appendix shows some summary statistics of the labor market outcomes analyzed in this section. Table A8 in the Appendix shows that survey response is unrelated to being exposed to student instructors.

power: it could be that, even with our large sample size, some of the student instructor effects we are trying to measure are simply not large enough to be detected by any *single* outcome we observe. The second is a problem of inference: statistically testing for student instructor effects on 14 different outcomes potentially leads some of these tests to incorrectly reject their null hypothesis. Both these problems are addressed in the context of treatment effects of early childhood interventions by Anderson (2008), on which we base this subsection.

Both problems of power and inference in our analyses can be addressed using a summary index test corrected for familywise error rates and false discovery rates. To do this, we first construct summary indices of the outcomes variables we believe to be capturing the same "core outcome" for students. Our indices measure: *instructor rating* (combining all instructor related course evaluation items), *course rating* (combining course evaluation items "course evaluation," "tutorial group functioning" and "material helpfulness"), *subsequent earnings* (combining log of first earning and log of current earnings), and *reported satisfaction* (combining study and job satisfaction).[20] Each summary index is a weighted average of its items. The weight is the inverse of the items' correlation to one another, maximizing the independent information captured in the index. The indices are normalized to have zero mean and a standard deviation equal to the standard deviation of the base group (in this case, the outcomes of students taught by non-student instructors). This normalization eases the interpretation of the results.

We then estimate the effect of student instructors on these four summary indices controlling for the same covariates we included in the previous regressions. Finally, we correct for familywise error and false discovery using the step-down procedure of Benjamini, Krieger, and Yekutieli (2006) as implemented by Anderson (2008). This last procedure produces

[20] We chose to not include student effort in the indices since it is not clear whether this is a desirable student outcome and the index construction requires us to determine this *a priori*. We do not construct an index for job search length since this outcome is only measured through one item in our labor market survey data.

adjusted p-values (called sharpened q-values) which should be used for inference instead of the traditional p-values.

Table 9 presents the results of the process described above. The point estimates show that students exposed to student instructors generally rate their instructors and the courses they teach worse and have less earnings after graduation, consistent with the overall message of our analyses in Sections 4.1 and 4.2. However, the original p-values show that only the effect on instructor ratings is marginally statistically significant. More importantly, when we correctly base our inference on the sharpened q-values, all results become statistically insignificant. We therefore conclude that our analyses in Section 4 fails to provide any strong evidence that student instructors affect their students in any way we can capture with either our course evaluation measures or our post-graduation outcome measures.[21]

5 Conclusion

This paper investigates the effectiveness of a frequently used, yet understudied, input in university education—the student instructor. We show that being taught by a student instructor compared to a more senior instructor has only a tiny negative effect on students' current grades. These effects are not cumulative, nor are they persistent. The small effect sizes do not appear to be driven by differences in grading standards between student and non-student instructors. We find weak evidence that student instructors are related to lower student ratings of the knowledge of their instructors, of the course material, and of the course itself. When looking at students' outcomes after graduation, we find no evidence that student instructors are detrimental to students' job search length, earnings, or job satisfaction after they left the university.

[21] As an additional check, we also jointly correct all the main estimates in this paper for familywise error rates and false discovery rates. The results, presented in Table A9 in the Appendix, show that once these corrections are made, we can only provide evidence that student instructors negatively affect their students' rating of the course material.

The lack of any sizable effect on the wide range of academic and labor market outcomes we looked at is surprising. These findings hint to the subtle nature of teacher quality and the complexity of the learning process in higher education. The results could also be driven by at least three mechanisms. Maybe student instructors compensate for their lack of knowledge and experience by being better able to relate the students. The fact that many of them have the course material fresh in their heads potentially makes them better equipped at explaining it. Or it could be that senior staff provide less effort in teaching, nullifying the possible returns to their qualifications and narrowing the gap between them and student instructors. Finally, it could simply be that what makes a good teacher is really unrelated to whether an instructor is a student teacher. Differentiating between these mechanisms is crucial for the proper design of staffing policies and teaching incentives, and we hope to develop this research avenue in the future.

Our results can inform university policy by showing that universities can liberate financial resources by expanding the use of student instructors at almost no cost in terms of students' achievement. If enough students are willing and able to teach tutorials, they could also be used to lighten the teaching load of senior staff.

References

Anderson, M. L. (2012). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), 1481-1495.

Baltagi, B., (2005). *Econometric Analysis of Panel Data*. 3rd ed. West Sussex, England: John Wiley & Sons Ltd.

Benjamini, Y., Krieger, A.M. and Yekutieli, D., 2006. Adaptive Linear Step-up Procedures that Control the False Discovery Rate. *Biometrika*, 93(3), 491-507.

Bettinger, E. P., Long, B. T., & Taylor, E. S. (2016). When Inputs are Outputs: The case of Graduate Student Instructors. *Economics of Education Review*, *52*, 63-76.

Bettinger, E. P., & Long, B. T. (2010). Does Cheaper Mean Better? The Impact of Using Adjunct Instructors on Student Outcomes. *The Review of Economics and Statistics*, *92*(3), 598-613.

Borjas, G. J. (2000). Foreign-Born Teaching Assistants and the Academic Performance of Undergraduates. *American Economic Review*, *90*(2), 355-359.

Carrell, S. E., and West, J.E (2010). Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors. *Journal of Political Economy* (118, no. 3).

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *The American Economic Review*, *104*(9), 2593-2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, *104*(9), 2633-2679.

Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *The American Economic Review*, *97*(2), 31-47.

De Vlieger, P., Jacob, B., & Stange, K. (2017). Measuring Instructor Effectiveness in Higher Education. In *Productivity in Higher Education*. University of Chicago Press.

Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2014). A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom. *The American Economic Review*, *104*(8), 2567-2591.

Feld, J., Salamanca, N, and Hamermesh, D.S. (2016). Endophilia or Exophobia: Beyond Discrimination. *The Economic Journal*, 126(594), 1503–1527.

Feld, J., & Zölitz, U. (2017). Understanding Peer Effects: On the Nature, Estimation and Channels of Peer Effects. *Journal of Labor Economics*, 35(2), 387-428.

Figlio, D. N., Schapiro, M. O., & Soter, K. B. (2015). Are Tenure Track Professors Better Teachers? *Review of Economics and Statistics*, *97*(4), 715-724.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica*, *77*(3), 953-973.

Fleisher, B., Hashimoto, M., & Weinberg, B. A. (2002). Foreign GTAs can be Effective Teachers of Economics. *The Journal of Economic Education*, *33*(4), 299-325.

Harris, D. N., & Sass, T. R. (2011). Teacher Training, Teacher Quality and Student Achievement. *Journal of Public Economics*, 95(7), 798-812.

Hoffmann, F., & Oreopoulos, P. (2009). Professor Qualities and Student Achievement. *The Review of Economics and Statistics*, *91*(1), 83-92.

Jepsen, C., & Rivkin, S. (2009). Class Size Reduction and Student Achievement the Potential Tradeoff Between Teacher Quality and Class Size. *Journal of Human Resources*, *44*(1), 223-250.Lusher, Lester, Doug Campbell, & Scott Carrell. *TAs Like Me: Racial Interactions between Graduate Teaching Assistants and Undergraduates*. No. w21568. National Bureau of Economic Research, 2015.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

Zölitz, U. & Feld, J. (2017). The Effect of Peer Gender on Major Choice and Occupational Segregation. Unpublished manuscript.

TABLES

Table 1. Characteristics of all Courses and Courses that use at least one Student and one Non-student Instructor (Sample Courses)

	All Courses (N = 575)	Sample Courses (N = 173)		
	Mean	Mean	Diff. in Means	[p-value]
Student Instructor	0.14	0.43	-0.29	[0.000]
PhD Student Instructor	0.28	0.19	0.09	[0.000]
Senior Instructor	0.57	0.37	0.20	[0.000]
Student GPA at Signup	6.90	6.87	0.03	[0.285]
Final Course Grade	6.93	6.78	0.15	[0.009]
Mathematical Course	0.24	0.27	-0.03	[0.441]
First-year Course	0.13	0.28	-0.15	[0.000]
Bachelor Student	0.66	0.83	-0.17	[0.000]
No. of Tutorial Groups	9.16	14.65	-5.49	[0.000]
No. of Students	114.87	188.41	-73.54	[0.000]
No. of Students per Tutorial Group	12.28	12.67	-0.39	[0.003]
No. of Instructors	3.57	5.43	-1.86	[0.000]

This table is based on data comprising 61,733 course final grades from 12,609 students who took 144 different courses, taught by 578 instructors over 23 teaching periods between 2009 and 2014. The difference in means tests is performed using an unpaired sample t-test with unequal variances.

Table 2. Summary Statistics for Instructors, Students, and Tutorial groups

Panel A: Instructors (N = 434)				
	Mean	S.D.	Min	Max
Student Instructor	0.50	0.50	0	1
PhD Student Instructor	0.25	0.44	0	1
Senior Instructor	0.24	0.43	0	1
Female Instructor	0.38	0.49	0	1
Dutch Instructor	0.30	0.46	0	1
German Instructor	0.43	0.50	0	1
No. of Courses	3.96	5.15	1	44
No. of Tutorial Groups	10.39	14.32	1	113
No. of Students	131.12	181.09	10	1444

Panel B: Students (N = 6,649)				
	Mean	S.D.	Min	Max
Female Student	0.39	0.49	0	1
Dutch Student	0.29	0.46	0	1
German Student	0.51	0.50	0	1
Bachelor Student	0.81	0.39	0	1
Final Course Grade	6.70	1.39	1	9.5
GPA	6.63	1.32	1.33	10
No. of Courses	7.61	4.87	1	23
Age	20.60	2.13	16.25	41.25

(continued on next page)

(Table 2 continued)

Panel C: Tutorial groups (N = 2,534)

	Mean	S.D.	Min	Max
Student Instructor	0.42	0.49	0	1
PhD Student Instructors	0.18	0.39	0	1
Senior Instructor	0.40	0.49	0	1
Mathematical Course	0.30	0.46	0	1
First-year Course	0.41	0.49	0	1
Female Instructor	0.38	0.48	0	1
Dutch Instructor	0.37	0.48	0	1
German Instructor	0.32	0.47	0	1
Other Nationality Instructor	0.30	0.46	0	1
No. of Students	12.86	1.58	1	16
Student GPA	6.92	0.52	4.08	9

This table is based on our estimation sample comprising 28,203 course final grades from 6,649 students who took 173 different courses, taught by 434 instructors over 23 teaching periods between 2009 and 2014.

Table 3. Balancing Test of Student Instructors on Pre-Assignment Characteristics

Dep. Variable:	Student GPA	Female Student	Student Age	Student ID
	(1)	(2)	(3)	(4)
Student Instructor	0.013	-0.011*	-0.031	-35.042
	(0.017)	(0.006)	(0.019)	(51.865)
PhD Student Instructor	0.015	-0.006	-0.029	-57.380
	(0.023)	(0.009)	(0.023)	(70.989)
F-Test p-value:	[0.711]	[0.209]	[0.227]	[0.673]
Fixed Effects:	✓	✓	✓	✓
R-squared	0.12	0.07	0.42	0.04
Observations	28,203	28,203	28,203	28,203
Instructors	434	434	434	434

*This table reports OLS coefficients of regressing student pre-treatment characteristics on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor). All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the teacher level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 4. The Effects of Student Instructor on Contemporaneous Grades

Dep. Variable: Std. Final Grade	(1)	(2)	(3)
Student Instructor	-0.018	-0.018	-0.023*
	(0.014)	(0.014)	(0.013)
Std. GPA			0.593***
			(0.010)
Student Characteristics:		✓	✓
Fixed Effects:	✓	✓	✓
R-Squared	0.18	0.21	0.51
Observations	28,203	28,203	28,203
Instructors	434	434	434

*This table reports OLS coefficients of regressing standardized (Std. Dev.=1) final course grades on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor) and student GPA before taking the course. Student characteristics include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 5. Heterogeneous Effects of Student Instructors by Course and Student Type

Dep. Variable:	First-Year Course		Mathematical Course		Lower Ability Students	
Std. Final Grade	Yes	No	Yes	No	Yes	No
	(1)	(2)	(3)	(4)	(5)	(6)
Student Instructor	-0.008	-0.033*	-0.001	-0.030*	-0.045**	-0.000
	(0.015)	(0.017)	(0.016)	(0.016)	(0.020)	(0.014)
Std. GPA	0.634***	0.536***	0.657***	0.559***	0.523***	0.658***
	(0.012)	(0.015)	(0.019)	(0.010)	(0.018)	(0.014)
Other covariates:	✓	✓	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓	✓	✓
R-Squared	0.58	0.44	0.55	0.48	0.41	0.41
Observations	12,170	16,033	8,568	19,635	13,708	14,495
Instructors	236	288	186	308	432	433

*This table reports OLS coefficients of regressing standardized (Std. Dev.=1) final course grades on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor) and student GPA before taking the course. High ability students are students with above-median GPA in each course. Mathematical courses are defined in the main text. First-year courses are courses exclusively given in the first year of bachelor programs. Other covariates include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 6. Effect of Student Instructor on Subsequent Average Grades

Dep. Variable: Std. GPA (after 2nd year)	(1)	(2)	(3)
Student Instructor (first year)	-0.005	-0.003	-0.006
	(0.018)	(0.018)	(0.016)
Std. GPA (first year)			0.415***
			(0.008)
Student Characteristics:		✓	✓
Fixed Effects:	✓	✓	✓
R-Squared	0.04	0.09	0.44
Observations	7,075	7,075	7,075
Instructors	196	196	196

*This table reports OLS coefficients of regressing standardized (Std. Dev.=1) student GPA after second year on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor) and student GPA. All independent variables refer to first-year courses. Student characteristics include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 7. Heterogeneous Effects by Instructor Ability

Dep. Variable: Std. Final Grade	(1)	(2)	(3)
Instructor Std. GPA	0.078		
	(0.059)		
High Ability Student Instructor		-0.015	
		(0.017)	
Low Ability Student Instructor		-0.008	
		(0.022)	
Quintiles of Student Instructor Ability:			
1st			-0.025
			(0.022)
2nd			-0.012
			(0.024)
3rd			-0.041
			(0.034)
4th			0.005
			(0.021)
5th			0.027
			(0.041)
Std. GPA	0.593***	0.593***	0.593***
	(0.010)	(0.010)	(0.010)
F-Test p-value:	[0.187]	[0.748]	[0.624]
Other Covariates:	✓	✓	✓
Fixed Effects:	✓	✓	✓
R-Squared	0.51	0.51	0.51
Observations	28,203	28,203	28,203
Instructors	434	434	434

*This table reports OLS coefficients of regressing standardized (Std. Dev.=1) final course grades on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor) and student GPA before taking the course. Instructor ability indicators are standardized GPA of Student Instructors, a below- and above-median GPA division (median GPA = 7.76), and dummies for instructor GPA quintiles. Other covariates include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 8. Effect of Student Instructor on Student Post-Graduation Outcomes

Dep. Variable:	Unemp. after graduation:		Log earnings:		Satisfaction with:	
	None	Months	First	Current	Studies	Job
	(1)	(2)	(3)	(4)	(5)	(6)
Student Instructor	0.002	-0.075	-0.017	-0.019	0.025	-0.013
	(0.011)	(0.062)	(0.026)	(0.030)	(0.024)	(0.037)
Std. GPA	0.054***	-0.234***	0.080***	0.010	0.165***	0.118***
	(0.005)	(0.035)	(0.013)	(0.014)	(0.013)	(0.019)
Other covariates:	✓	✓	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓	✓	✓
R-Squared	0.13	0.08	0.14	0.14	0.10	0.09
Observations	11,539	8,793	7,868	9,518	11,539	8,358
Instructors	413	411	411	411	413	412

*This table reports OLS coefficients of regressing students' labor market outcomes from a post-graduation survey on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor) and student GPA before taking the course. Unemployment after graduation is measured as a dummy demarking having a job lined up before graduating (Column 1) and the median number of months of unemployment from a 6-category measure capped at 12 months (Column 2). Column 2 excludes those who have a job lined up after graduation and includes a dummy for "I did not (yet) start working after graduation." Earnings are measured in annualized thousands of euros. Satisfaction variables are measured from 1 to 10 and increasing in satisfaction. Other covariates include student gender and nationality, a cubic polynomial for student age, and a dummy for whether the survey was conducted by phone (vs online). All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 9. Effect of Student Instructor on Summary Indices of Student Outcomes

Dep. Variable:	Summary index for:			
	Instructor rating	Course rating	Subsequent earnings	Reported Satisfaction
	(1)	(2)	(3)	(4)
Student Instructor	-0.170	-0.080	-0.014	0.010
	(0.089)	(0.051)	(0.025)	(0.021)
Original p-values	[0.055]	[0.116]	[0.628]	[0.566]
Sharpened q-values	[0.283]	[0.283]	[0.458]	[0.458]
Other covariates:	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓
R-Squared	0.18	0.19	0.15	0.09
Observations	10,026	10,717	9,558	11,539
Instructors	391	428	411	413

*This table reports OLS coefficients on summary indices and "sharpened" two-stage q-values (Benjamini, Krieger, and Yekutieli, 2006) which correct for multiple testing as described in Anderson (2008). The construction from the summary indices are explained in the main text. The regressors include student instructor and PhD student instructor (unreported) dummy variables (the base group is senior instructor) and student GPA (unreported) before taking the course and adjusting the values. Other covariates include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1 based on the sharpened q-value calculations.*

FIGURES

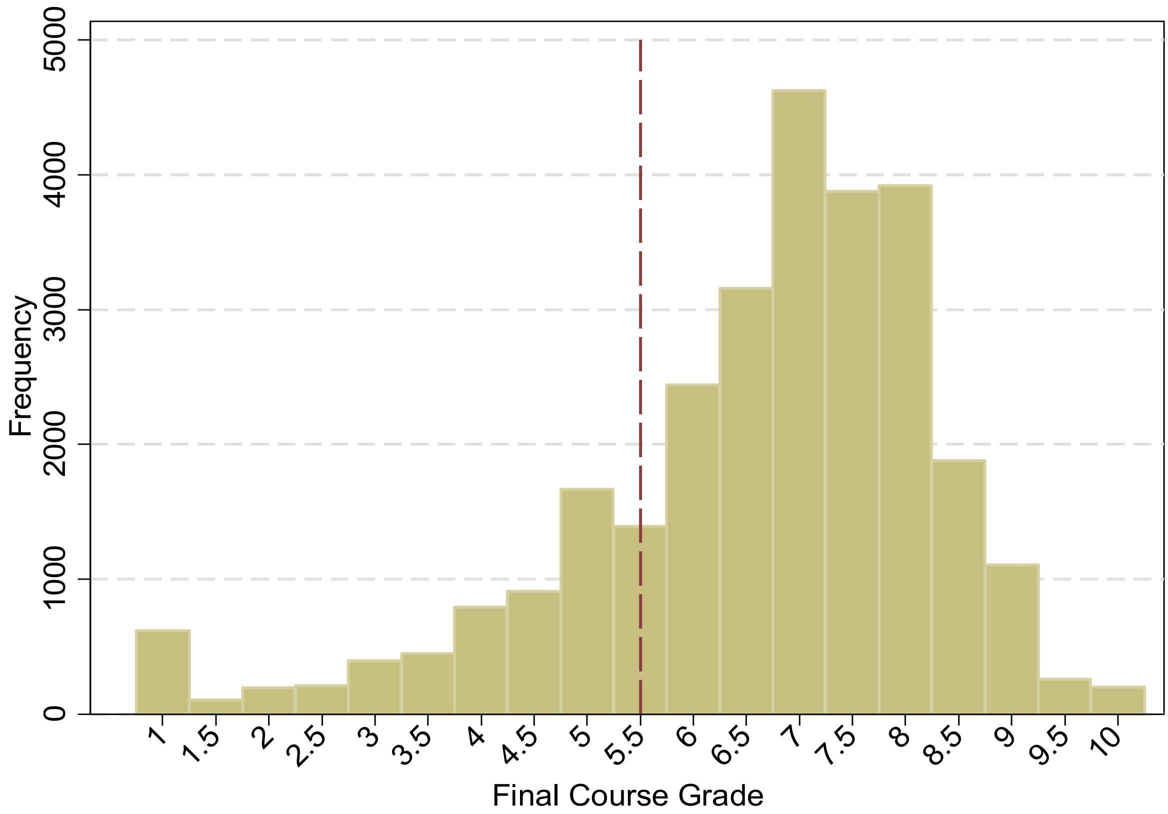

Fig. 1. Distribution of Course Final Grades

Note: This figure is based on the estimation sample. The vertical line at 5.5 shows the lowest possible passing grade.

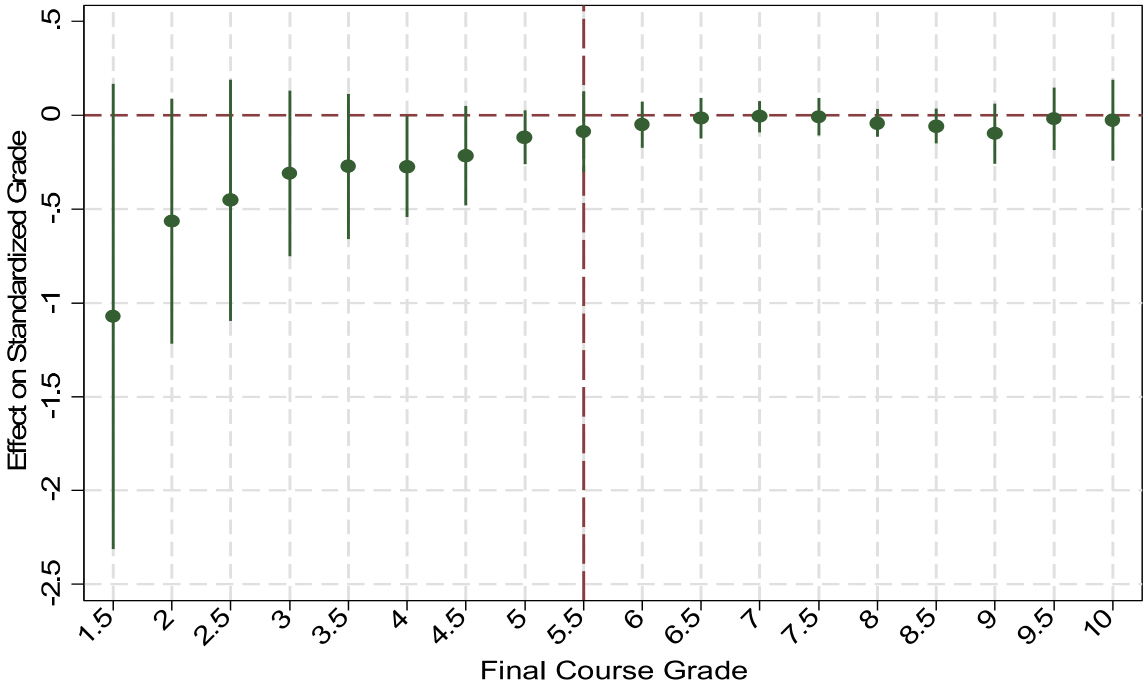

Fig. 2. Quantile Treatment Effects of Student Instructors

Note: This figure is based on regression estimates shown in Table A2. The doted vertical line at 5.5 shows the lowest possible passing grade. The solid vertical lines show 95 percent confidence intervals.

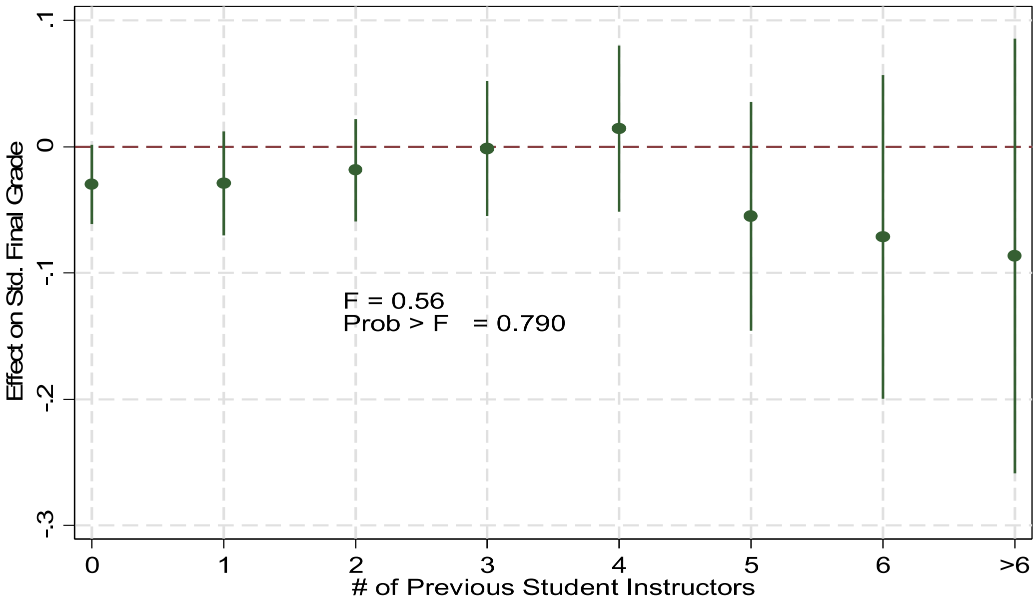

Fig. 3. Cumulative Effect of Student Instructor on Grades

Note: This figure is based on regression estimates shown in Table A3. Vertical lines show 95 percent confidence intervals.

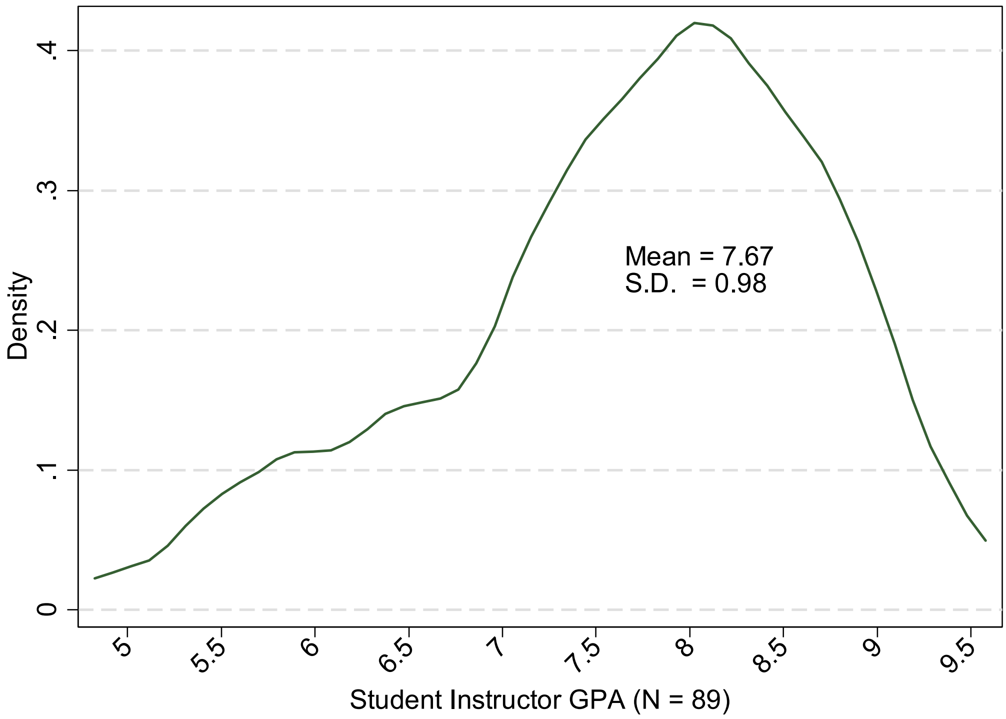

Fig. 4. Distribution of Student Instructor Grade Point Average (GPA)

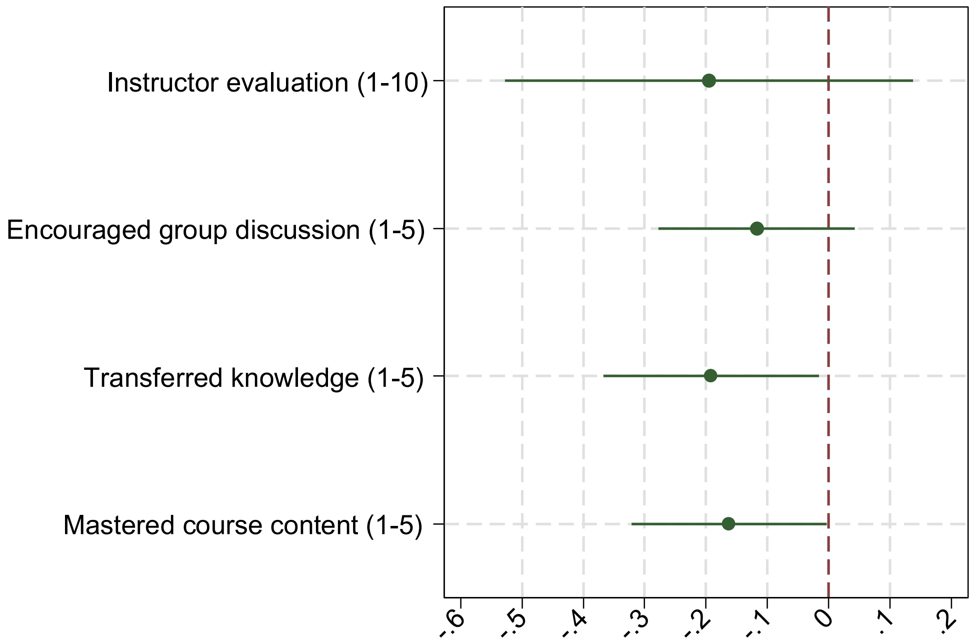

Fig. 5. Effect of Student Instructor on Instructor-Related Evaluation Outcomes

Note: This figure is based on regression estimates shown in Table A6. Horizontal lines show 95 percent confidence interval.

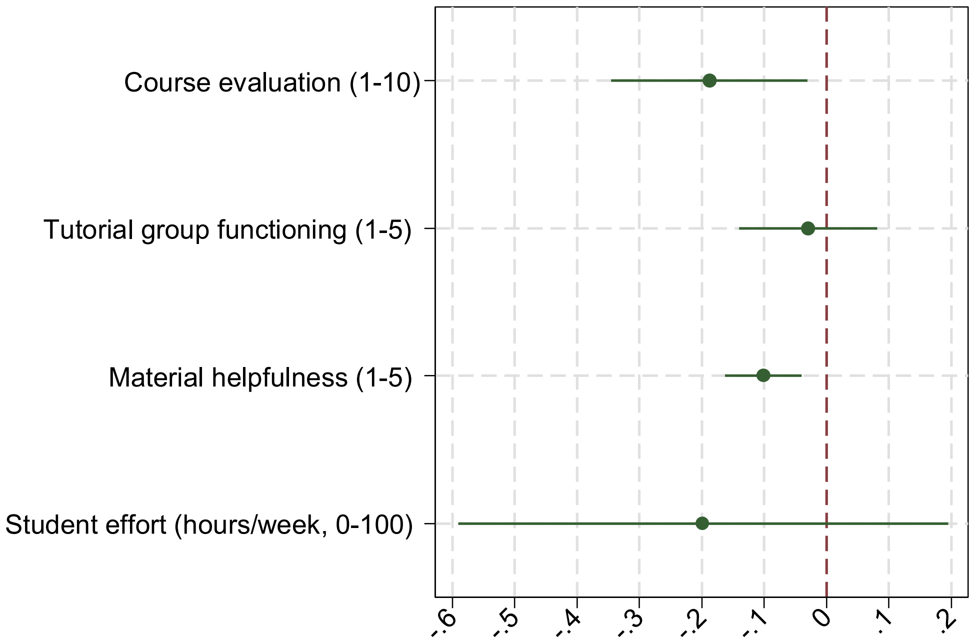

Fig. 6. Effect of Student Instructor on Other Evaluation Outcomes

Note: This figure is based on regression estimates shown in Table A6. Horizontal lines show 95 percent confidence interval.

APPENDIX

A1 Data Restrictions

Below we list the observations we exclude from our estimation sample because they represent exceptions from the standard tutorial group assignment procedure at the SBE.

- We exclude eight courses in which the course coordinator or other education staff actively influenced the tutorial group composition. One course coordinator, for example, requested to balance student gender across tutorial groups. The SBE scheduling department informed us about these courses.
- We exclude 21 tutorial groups from the analysis that consisted mainly of students who registered late to the course. Before April 2014, SBE reserved one or two slots per tutorial group for students that registered late. In exceptional cases where the number of late registration students substantially exceeded the number of empty spots, new tutorial groups were created that mainly consist of late registering students. SBE abolished the late registration policy in April 2014.
- We exclude 46 repeater tutorial groups from the analysis. One course coordinator explicitly requests to assign repeater students who failed his courses in the previous year to special repeater tutorial groups.
- We exclude 17 tutorial groups that consist mainly of MARBLE (Maastricht Research Based Learning program) students. For some courses, MARBLE students are assigned together to separate tutorial groups with more experienced teacher.
- We exclude 95 part-time MBA students, since these students are typically scheduled for special evening classes with only part-time students.

- We exclude 4,274 student-year observations for students who were repeating courses. These students follow a different attendance criteria and are graded under different standards.
- We exclude all observations of the first year and the first period students are observed. For these observations, we have no measure of previous performance of the student at the SBE, an essential covariate in our analyses.
- We exclude all observations from the first teaching period of 2009—the first period in our dataset—for the same reasons outlined above
- We exclude 1,229 tutorial groups which take place after 6:30 pm since before Fall 2015 students had the option to opt out of evening education, which makes the student assignment to these tutorials potentially non-random.

Table A1. Use of Student Instructors in OECD Countries

Country	Student instructors used in a typical university?
AUSTRALIA	Yes
AUSTRIA	Yes
BELGIUM	No
CANADA	Yes
CHILE	Yes
CZECH REPUBLIC	No
DENMARK	Yes
ESTONIA	Yes
FINLAND	Yes
FRANCE	Yes
GERMANY	Yes
GREECE	No
HUNGARY	Yes
ICELAND	Yes
IRELAND	Yes
ISRAEL	Yes
ITALY	Yes
JAPAN	Yes
KOREA	Yes
LATVIA	No
LUXEMBOURG	No
MEXICO	Yes
NETHERLANDS	Yes
NEW ZEALAND	Yes
NORWAY	Yes
POLAND	Yes
PORTUGAL	No
SLOVAK REPUBLIC	Yes
SLOVENIA	Yes
SPAIN	No
SWEDEN	No
SWITZERLAND	Yes
TURKEY	Yes
UNITED KINGDOM	No
UNITED STATES	Yes

We collected this information by contacting people with experience in higher education institutions in these countries by email. We asked: "Student instructors can be bachelor or master students that teach at university, typically in small group teaching like tutorials, exercises or lab sessions. Are student instructors used in a typical university in <<name of the country>>". The answer to this question of course depends on the specific experiences of the respondents. While the answer for any individual country might be wrong, the overall picture that emerges is unambiguous: student instructors are used in many OECD countries.

Table A2. Quantile Treatment Effects of Student Instructors

Dep. Variable: Std. Final Grade	Unconditional Quantile Treatment Effect at Final grade:								
Final Grade (quantile) =	*1.5*	*2*	*2.5*	*3*	*3.5*	*4*	*4.5*	*5*	*5.5*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Student Instructor	-1.073*	-0.564*	-0.452	-0.310	-0.272	-0.273**	-0.216	-0.118	-0.086
	(0.631)	(0.332)	(0.328)	(0.224)	(0.197)	(0.137)	(0.135)	(0.074)	(0.109)
Std. GPA	6.929***	3.643***	4.586***	4.375***	3.841***	3.914***	4.294***	2.342***	4.213***
	(0.612)	(0.322)	(0.327)	(0.212)	(0.186)	(0.136)	(0.133)	(0.073)	(0.097)
Other covariates:	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓	✓	✓	✓	✓	✓
R-Squared	0.20	0.20	0.19	0.20	0.20	0.24	0.28	0.28	0.37
Observations	28,203	28,203	28,203	28,203	28,203	28,203	28,203	28,203	28,203
Instructors	434	434	434	434	434	434	434	434	434

(continued on next page)

(Table A2 continued)

Dep. Variable: Std. Final Grade	Unconditional Quantile Treatment Effect at Final grade:								
Final Grade (quantile) =	*6*	*6.5*	*7*	*7.5*	*8*	*8.5*	*9*	*9.5*	*10*
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Student Instructor	-0.049	-0.016	-0.008	-0.009	-0.042	-0.057	-0.097	-0.020	-0.026
	(0.062)	(0.055)	(0.043)	(0.051)	(0.038)	(0.048)	(0.081)	(0.085)	(0.110)
Std. GPA	2.401***	2.092***	1.484***	1.770***	0.900***	1.021***	1.730***	1.215***	1.576***
	(0.055)	(0.043)	(0.025)	(0.029)	(0.031)	(0.050)	(0.085)	(0.177)	(0.230)
Other covariates:	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓	✓	✓	✓	✓	✓
R-Squared	0.37	0.42	0.41	0.41	0.26	0.19	0.19	0.10	0.10
Observations	28,203	28,203	28,203	28,203	28,203	28,203	28,203	28,203	28,203
Instructors	434	434	434	434	434	434	434	434	434

*This table reports OLS coefficients of regressing the recentered influence function (RIF) of standardized final course grades on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor) and student GPA before taking the course. The RIF parallels Firpo, Fortin and Lemieux (2009) and is calculated at the corresponding quantile of every point in our discrete grade distribution. Other covariates include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table A3. Cumulative Effect of Student Instructor on Grades

Dep. Variable: Std. Final Grade	Main effects:	Interactions with Student Instructor:
Student Instructor	-0.030*	-
	(0.016)	
Std. GPA	0.590***	-
	(0.010)	
Previous student Instructors = 1	0.001	0.001
	(0.017)	(0.024)
Previous student Instructors = 2	0.056***	0.011
	(0.017)	(0.024)
Previous student Instructors = 3	0.073***	0.028
	(0.023)	(0.030)
Previous student Instructors = 4	0.050*	0.044
	(0.027)	(0.036)
Previous student Instructors = 5	0.096***	-0.025
	(0.035)	(0.048)
Previous student Instructors = 6	0.079*	-0.042
	(0.041)	(0.068)
Previous student Instructors > 6	0.014	-0.057
	(0.068)	(0.089)
F-Test interactions p-value:	[0.790]	
Other Covariates:	✓	
Fixed Effects:	✓	
R-Squared	0.51	
Observations	28,203	
Instructors	434	

*This table reports OLS coefficients of regressing standardized (Std. Dev.=1) final course grades on a student instructor and interacted with the number of student instructors each student has been exposed to in the past. Other covariates include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table A4. Effect of Student Instructor on Course Evaluation Survey Response

Panel A: Course Evaluations - Instructor Performance				
Dep. Var. Responded to item:	Instructor Evaluation	Encouraged Group Discussion	Transferred Knowledge	Mastered Course Content
	(1)	(2)	(3)	(4)
Student Instructor	0.011	0.010	0.011	0.010
	(0.009)	(0.009)	(0.009)	(0.009)
Std. GPA	0.043***	0.042***	0.042***	0.042***
	(0.003)	(0.003)	(0.003)	(0.003)
Other Covariates:	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓
R-squared	0.14	0.15	0.15	0.15
Observations	28,203	28,203	28,203	28,203
Instructors	434	434	434	434

Panel B: Course Evaluations - Other Outcomes				
Dep. Var. Responded to item:	Course Evaluation	Tutorial Group Functioning	Material Helpfulness	Student Effort (hours/week)
	(1)	(2)	(3)	(4)
Student Instructor	0.012	0.011	0.012	0.005
	(0.009)	(0.009)	(0.009)	(0.008)
Std. GPA	0.047***	0.042***	0.037***	0.044***
	(0.003)	(0.003)	(0.003)	(0.003)
Other Covariates:	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓
R-squared	0.11	0.14	0.17	0.14
Observations	28,203	28,203	28,203	28,203
Instructors	434	434	434	434

*This table reports OLS coefficients of regressing survey response dummies for each survey variable on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor) and student GPA before taking the course. The definition of all the survey response variables and their main summary statistics can be found in Table A4. Other covariates include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table A5. Summary Statistics and Question Description of Course Evaluations

			Summary Statistics:				
Item	Response Range	Text	Response Rate in sample	Completion rate if survey started	Mean	Min	Max
Instructor Evaluation	1-10	*Evaluate the overall functioning of your tutor in this course with a grade: (1 = very bad, 6 =sufficient, 10 = very good).*	35.0%	91.7%	7.81	1	10
Encouraged Group Discussion	1-5	*The tutor encouraged all students to participate in the (tutorial) group discussions.*	35.2%	92.1%	3.64	1	5
Transferred Knowledge	1-5	*The tutor stimulated the transfer of what I learned in this course to other contexts.*	35.3%	92.5%	3.97	1	5
Mastered Course Content	1-5	*The tutor sufficiently mastered the course content.*	35.3%	92.6%	4.31	1	5
Course Evaluation	1-10	*Please give an overall grade for the quality of this course (1=very bad, 6=sufficient, 10=very good)?*	37.5%	98.2%	7.04	1	10
Tutorial Group Functioning	1-5	*My tutorial group has functioned well.*	35.3%	92.4%	3.95	1	5
Material Helpfulness	1-5	*The textbook, the reader and/or electronic resources helped me studying the subject matters of this course.*	32.4%	84.9%	3.64	1	5
Student Effort (hours/week)	0-100	*How many hours per week on the average (excluding contact hours) did you spend on self-study (presentations, cases, assignments, studying literature, etc)?*	32.6%	85.5%	13.93	0	60

Table A6. Effect of Student Instructor on Course Evaluation Outcomes

Panel A: Course Evaluations - Instructor Performance				
Dep. Var:	Instructor Evaluation	Encouraged Group Discussion	Transferred Knowledge	Mastered Course Content
	(1)	(2)	(3)	(4)
Student Instructor	-0.195	-0.117	-0.192**	-0.162**
	(0.169)	(0.081)	(0.089)	(0.081)
Std. GPA	-0.140***	-0.141***	-0.096***	-0.034***
	(0.030)	(0.017)	(0.015)	(0.012)
Other Covariates:	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓
R-squared	0.17	0.15	0.17	0.16
Observations	9,870	9,917	9,955	9,967
Instructors	391	387	387	387

Panel B: Course Evaluations - Other Outcomes				
Dep. Var:	Course Evaluation	Tutorial Group Functioning	Material Helpfulness	Student Effort (hours/week)
	(1)	(2)	(3)	(4)
Student Instructor	-0.188**	-0.029	-0.101***	-0.198
	(0.080)	(0.056)	(0.031)	(0.200)
Std. GPA	-0.196***	-0.095***	-0.047***	-0.215*
	(0.023)	(0.013)	(0.015)	(0.111)
Other Covariates:	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓
R-squared	0.21	0.16	0.20	0.21
Observations	10,577	9,952	9,135	9,206
Instructors	428	391	385	391

*This table reports OLS coefficients of regressing course survey variables on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor), student GPA before taking the course, and student course final grade. The definition of all the survey response variables and their main summary statistics can be found in the Appendix. Other covariates include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table A7. Sample Comparison and Summary Statistics of Students' Labor Market Outcomes

Panel A: Post-Graduation Survey Sample Comparison			
	BA Students (N = 5,504)	Survey Respondents (N = 1,618)	
	Mean	Mean	Diff. in Means
Share of student instructors	0.25	0.29	-0.04
Share of PhD student instructors	0.24	0.23	0.01
Female Student	0.38	0.38	0.00
Dutch Student	0.27	0.29	-0.02
German Student	0.53	0.58	-0.05
Other Nationality Student	0.19	0.13	0.06
Bachelor Student	0.97	0.99	-0.02
GPA	6.43	6.86	-0.43
No. of Courses	7.99	11.31	-3.32
Age	20.08	20.23	-0.15

Panel B: Post-Graduation Survey Summary Statistics			
	Obs.	Mean	S.D.
Job Search Length After Graduation:			
Job Lined up Already	1,197	0.44	0.50
0-1 months	1,197	0.15	0.36
1-2 months	1,197	0.13	0.33
3-4 months	1,197	0.10	0.30
4-6 months	1,197	0.08	0.27
6-12 months	1,197	0.04	0.20
More than 12 Months	1,197	0.02	0.13
Did not (yet) Start Working	1,197	0.05	0.21
First Job Earnings ('000 euros yearly)	1,077	42.06	39.77
Current Earnings ('000 euros yearly)	1,307	45.92	36.82
Satisfaction with Studies (1-10)	1,618	8.10	1.17
Satisfaction with Job (1-10)	1,145	8.10	1.45

This table compares the sample of all 5,504 bachelor students between 2009 and 2014 and the subsample of 1,618 students who responded to our graduate survey.

Table A8. Effect of Student Instructor on Post-Graduation Survey Response

Dep. Var. Responded to item:	Unemployment After Grad.	First Earnings After Grad.	Current Earnings	Study Satisfaction	Job Satisfaction
	(1)	(2)	(3)	(4)	(5)
Student Instructor	-0.000	-0.003	-0.001	-0.003	0.001
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Std. GPA	0.040***	0.033***	0.050***	0.065***	0.043***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
Other Covariates:	✓	✓	✓	✓	✓
Fixed Effects:	✓	✓	✓	✓	✓
R-squared	0.24	0.21	0.19	0.29	0.22
Observations	28,203	28,203	28,203	28,203	28,203
Instructors	434	434	434	434	434

*This table reports OLS coefficients of regressing survey response dummies for each post-graduation survey variable on a student instructor and a PhD student instructor (unreported) dummy variable (the base group is senior instructor) and student GPA before taking the course. The definition of all the survey response variables and their main summary statistics can be found in the Appendix. Other covariates include student gender and nationality and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table A9. Familywise Error Rate and False Discovery Rate Correction for All Results

Panel A: Student Grades

Dep. Variable:	Concurrent Grades (1)	Subsequent Grades (2)
Student Instructor Effect	-0.023	-0.006
	(0.013)	(0.016)
Original p-values	[0.066]	[0.704]
Sharpened q-values	[0.189]	[0.677]

Panel B: Course Evaluation Outcomes

	Instructor Evaluation (1)	Encouraged Group Discussion (2)	Transferred Knowledge (3)	Mastered Course Content (4)	Course Evaluation (5)	Tutorial Group Functioning (6)	Material Helpfulness (7)	Student Effort (hours/week) (8)
Student Instructor Effect	-0.210	-0.124	-0.202	-0.168	-0.198	-0.035	-0.108**	-0.243
	(0.168)	(0.081)	(0.089)	(0.080)	(0.082)	(0.056)	(0.032)	(0.197)
Original p-values	[0.210]	[0.126]	[0.023]	[0.037]	[0.016]	[0.526]	[0.001]	[0.217]
Sharpened q-values	[0.384]	[0.301]	[0.130]	[0.140]	[0.130]	[0.677]	[0.017]	[0.384]

(continued on next page)

(Table A9 continued)

Panel C: Student Labor Market Outcomes						
	Months Unemp. After Grad.	Job Lined up After Grad.	Log First Earnings	Log Current Earnings	Study Satisfaction	Job Satisfaction
	(1)	(2)	(3)	(4)	(5)	(6)
Student Instructor Effect	-0.075	0.002	-0.018	-0.017	0.025	-0.013
	(0.062)	(0.011)	(0.026)	(0.030)	(0.024)	(0.037)
Original p-values	[0.227]	[0.854]	[0.488]	[0.568]	[0.294]	[0.713]
Sharpened q-values	[0.384]	[0.677]	[0.677]	[0.677]	[0.478]	[0.677]

*This table reports the original coefficients of all the main outcomes in the paper, together with their original p-values and their corresponding "sharpened" two-stage q-values (Benjamini, Krieger, and Yekutieli, 2006) corrected for multiple testing using the procedure reported in Anderson (2008). Each regression includes a PhD student instructor (unreported) dummy and the base group is senior instructor. Other characteristics include student GPA before taking the course, instructor gender and nationality, student gender and nationality, and a cubic polynomial for student age. All regressions condition on time-of-day and day-of-week fixed effects as well as course & other course combination fixed effects. Robust standard errors clustered at the instructor level in parentheses. Adjusted significance levels denoted by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$ based on the sharpened q-values.*

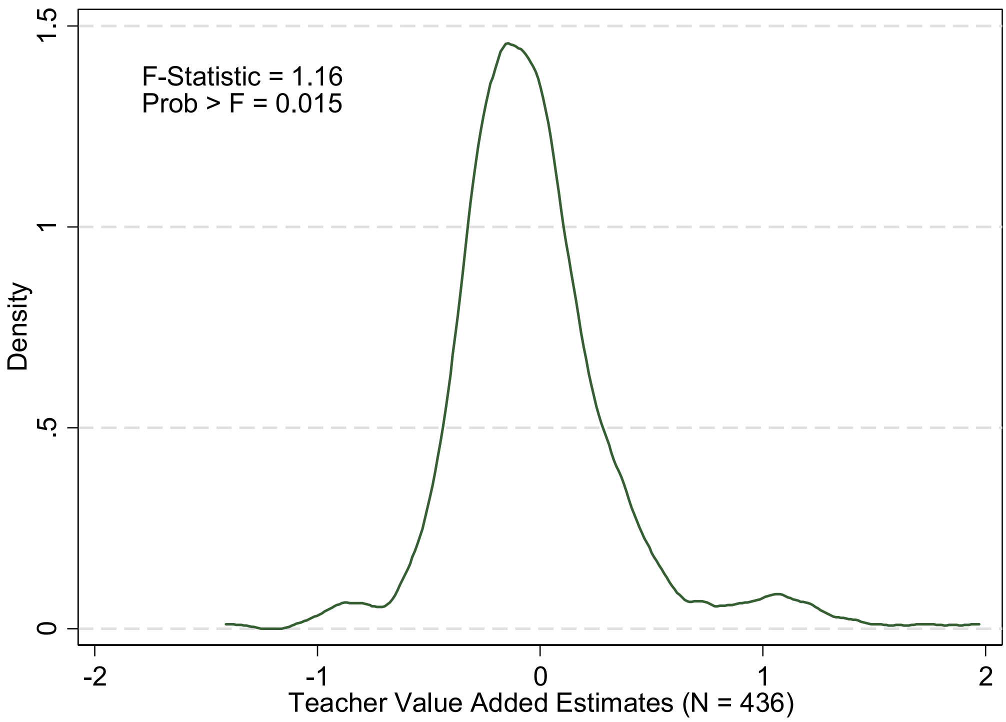

Fig. A1. Distribution of Instructor Effects in Estimation Sample

Note: This figure plots the instructor coefficients from a version of the model in Equation (1) where we replace instructor characteristics with instructor fixed effects.

THE UNIVERSITY OF MELBOURNE

MELBOURNE INSTITUTE